



Early Journal Content on JSTOR, Free to Anyone in the World

This article is one of nearly 500,000 scholarly works digitized and made freely available to everyone in the world by JSTOR.

Known as the Early Journal Content, this set of works include research articles, news, letters, and other writings published in more than 200 of the oldest leading academic journals. The works date from the mid-seventeenth to the early twentieth centuries.

We encourage people to read and share the Early Journal Content openly and to tell others that this resource exists. People may post this content online or redistribute in any way for non-commercial purposes.

Read more about Early Journal Content at <http://about.jstor.org/participate-jstor/individuals/early-journal-content>.

JSTOR is a digital library of academic journals, books, and primary source objects. JSTOR helps people discover, use, and build upon a wide range of content through a powerful research and teaching platform, and preserves this content for future generations. JSTOR is part of ITHAKA, a not-for-profit organization that also includes Ithaka S+R and Portico. For more information about JSTOR, please contact support@jstor.org.

A COEFFICIENT OF INDIVIDUAL PREPOTENCY FOR STUDENTS OF HEREDITY

DR. J. ARTHUR HARRIS

STATION FOR EXPERIMENTAL EVOLUTION, COLD SPRING HARBOR, N. Y.

I. THE CONCEPTION OF PREPOTENCY

THE term prepotency conveys the general idea that certain individuals "are particularly apt to impress their personal characters upon their offspring." Like most terms of general biology it has been applied in several different connections. The much-needed threshing over of the literature to separate the few measures of wheat from the stacks of straw and weeds falls outside the scope of this note.

One may follow Vilmorin, Hallett, Hays and many other noted breeders in the recognition of the practical importance of the fact that two individuals may be externally exactly alike and yet produce quite dissimilar offspring, without pledging himself to any of the theories of heredity in support of which it is sometimes cited. The aim of the practical breeder is not to formulate or to test theories of heredity but to get a strain of wheat which will draw the maximum amount of flour from an acre of soil or a breed of beasts that will yield the largest net dividends in milk, eggs or steak. His problem is pre-eminently a practical one, and one of the greatest services the student of biology can render him is to provide the criteria which enable him to select as easily as possible the parents of a race which will meet his requirements.

The purpose of this note is to call the attention of students of heredity to certain formulæ¹ which may be of

¹ These formulæ have heretofore been used in anthropometric surveys in testing the divergence of the inhabitants of a restricted community from the population of the whole area under consideration. They are equally

service in estimating the desirability of individual parents.

By individual prepotency² we understand for present purposes the phenomenon of certain individuals, or pairs of individuals in bi-parental inheritance, being exceptional in their capacity for producing offspring of any given characteristic.

As used here the term prepotency is most general. It implies nothing concerning the somatic similarity³ of parent and offspring and is in no way dependent upon any theory of heredity. It merely expresses a fact well known to practical breeders for half a century.

By a coefficient of individual prepotency one understands a statistical constant which measures the degree of superiority (with respect to the capacity for the production of offspring of any desired type) of any single parent or pair of parents.

well adapted to determine the significance of the deviation of an individual family from its generation.

The fundamental papers are:

Pearson, K., "On some Properties of the Hypergeometrical Series, and on the Fitting of such Series to Observation Polygons in the Theory of Chance," *Phil. Mag.*, February, 1899, p. 239.

Pearson, K., "On the Curves which are most Suitable for Describing the Frequency of Random Samples of a Population," *Biometrika*, Vol. V, pp. 172-175, 1906.

Pearson, K., "Note on the Significant or Non-significant Character of a Sub-sample drawn from a Sample," *Biometrika*, Vol. V, pp. 181-183, 1906.

Pearson, K., "On a Coefficient of Class Heterogeneity or Divergence," *Biometrika*, Vol. V, pp. 198-203, 1906.

Tocher, J. F., "The Anthropometric Characteristics of the Inmates of Asylums in Scotland," *Biometrika*, Vol. V, pp. 315-318, 1907.

Tocher, J. F., "Pigmentation Survey of School Children in Scotland," *Biometrika*, Vol. VI, pp. 143-146, 162-164.

²The disadvantages of using a word which has been so variously employed as prepotency are offset by keeping the terminology as simple as possible. The important thing is to have all terms carefully defined and unambiguous wherever used.

³In any study of heredity the correlation between the somatic characters of the parents and their offspring must be taken into account. Prepotency may, however, be estimated solely from the germinal characters of the parents as expressed in their respective arrays of offspring.

II. THE MEASUREMENT OF INDIVIDUAL PREPOTENCY

However well one may know the somatic characters of an individual or however intimate his knowledge of its ancestry the ultimate test of its value as a starting point for a new race is the quality of its offspring. *The proof of the parent is its produce* has been recognized as valid by various breeders since the time of Louis Vilmorin, who separated the parent beets and judged them individually by their offspring. The "ear-to-row" test in corn breeding, Petkus von Lochow's row-tests in rye and Hays's "centgener power" all represent attempts by practical breeders to obtain measures of individual prepotency as the term is used here. Galton's study of the distribution of prepotency in horses falls in the same class.

The method of estimating prepotency directly from the mean value (*e. g.*, sugar content) of the offspring, or from the number of offspring surpassing a given standard (*e. g.*, a mile in 2:30, or better, on the track) has disadvantages which will be obvious to those acquainted with elementary statistics.

So far as I am aware the credit of first recognizing the need of taking into account both type and variability in the criterion by which the relative desirability of the individual parents should be judged is due to Waugh.

In discussing some results secured on experiments with peas he remarks:⁴

. . . There were, as always, some exceptional cases of individual vines which showed a marked ability to transmit their individual characters to their offspring. The selection of such prepotent plants is evidently an important matter in plant breeding. In order to exhibit this difference we have computed a coefficient of heredity for each parent and for each character under study.

Waugh's formula is

$$C = 1/\sigma D,$$

where

C = coefficient of individual heredity,

σ = standard deviation of offspring,

⁴ Ann. Rept. Mass. Ag. Exp. Sta., Vol. 21, p. 172, 1909.

D = difference between numerical value of the parent character and the mean of the same character in the offspring.

In a later report he makes use of this formula "in an attempt to answer the question whether prepotency is inherited or not."⁵

Now while Waugh deserves all credit for suggesting the need of a coefficient of individual prepotency, I think the formula he proposes can not be justified theoretically nor regarded as practically satisfactory.⁶

The requirements of a coefficient of individual prepotency are at least the following:

(a) The comparison must be made between the offspring families, not between the individual parent and its offspring.

(b) The comparison must be so drawn as to attach importance only to differences significantly greater than the probable errors of random sampling.

(c) The coefficient expressing prepotency should be relative, *i. e.*, it should be comparable from character to character.

Proposition (b) and (c) will be granted without argument. In justification of (a) it is only necessary to point out that from the standpoint of the man who wishes to decide which families to continue to propagate and which to burn, the ideal method is one which may be applied to the individuals of any one generation entirely independently of those of any other. Of course this is not to be interpreted as a recommendation that in the routine work of practical or experimental breeding only one generation should be considered. What is meant is that it is desirable to have formulæ which permit of a consideration of prepotency on the data of any (offspring) generation independently. Such a formula does not preclude or render inadvisable the study of many ascendant generations.

⁵ Waugh, *loc. cit.*, Vol. 22, pp. 172-175, 1910.

⁶ Indeed he himself has pointed out some of the difficulties and has suggested that a better formula might be found.

The necessity of dealing with each generation independently is also imposed by the possibility of a differentiation between any two generations due to purely environmental (meteorological or edaphic) influences. Taken as a whole the entire offspring generation may be superior or inferior to the parental generation; and this because of no hereditary influence of the parents at all, for all families may be raised or lowered proportionally. This fact vitiates at once any comparison between individual parents and individual offspring unless the type and variability of both parental and offspring generation are taken into account.

In the practical work of calculation two cases may be conveniently recognized: in the first, the character is capable of direct measurement on a quantitative scale, in the second, the character is not capable of direct measurement but the individuals may be grouped into satisfactorily defined classes. In the first case the means may be compared; in the second case the proportional frequencies of one class must be used.

*(a) Case of Characters Measurable on a
Quantitative Scale*

It is well known that the standard deviation of a mean is σ/\sqrt{N} and its probable error is $.67449 \sigma/\sqrt{N}$. Given two uncorrelated means m and M , their difference and its probable error is given by

$$m - M \pm .67449 \sqrt{\frac{\sigma^2}{n} + \frac{\Sigma^2}{N}},$$

where σ and Σ and n and N represent the standard deviations and the numbers of individuals involved in the series. Thus it is quite easy to test the significance of differences in means between any two samples, or families in our case. But with a large number of families the labor of a series of such comparisons is prohibitive. What we need is some easily calculated criterion of the biological significance of the deviation of the mean of an

individual family from the mean of the population to which it belongs.

Suppose a population composed of N individuals with a mean of M and a variability of Σ is due to P parents. Now if this population be divided into two random samples of n and N' individuals, m and M' means, and σ and Σ' variabilities the differences in their means will be

$$(m - M') \pm .67449 \sqrt{\frac{\sigma^2}{n} + \frac{\Sigma'^2}{N'}}.$$

But Pearson has shown that the difference between the mean of a sub-sample m which in our case may represent the offspring of a single parent (or pair of parents) and the population mean M is not given by the preceding formula since n is included in N . The formula for such a case as this he has shown to be

$$(m - M) \pm .67449 \sqrt{\frac{\Sigma^2}{N} + \frac{\sigma^2}{n} \left(1 - \frac{2n}{N}\right) - \frac{n(M - m)^2}{N(N - n)}}.$$

This is the formula which we are seeking, the probable error of the difference between the mean for any family and that for the whole population. By calculating $(m - M)/E_{(m-M)}$ for every family we should have a criterion of its superiority or inferiority—the individual prepotency of the parent in question—relative to the average condition in the series to which it belongs.

Tocher has pointed out advantages in using $(m - M)/\sigma_{(m-M)}$ instead of $(m - M)/E_{(m-M)}$, but this is merely a matter of convenience. The significance of the ratios can be tested by tables of the normal curve.

(b) *Case of Characters not Measurable on a
Quantitative Scale*

For characters not quantitatively measurable two methods of treatment are available. The first consists in testing the divergence of a family from the general population on the basis of the relative frequency of a given character. The second consists in testing the deviation

of a family from the population with respect to the distribution of a character. At present the second of these methods seems of little practical importance for our purpose because of the relatively small numbers of individuals available in breeding experiments, even with plants, and because of the arithmetical routine.

Consider the first method. Let N be the number of individuals in a population due to P parents. Let X be a character common to all but appearing in different intensities (say from 0 development to the greatest possible intensity) in the several individuals, not measurable but capable of division into m classes. Let $s_1, s_2, s_3 \cdots s_m$ be the classes and $y_{s_1}, y_{s_2}, y_{s_3} \cdots y_{s_m}$ be the frequencies in the population as a whole. Now if a single family of n members be observed the probability of an individual belonging to any class, say s_2 , is $y_{s_2}/N = p$, while the probability of its not belonging to that class is $(1 - p) = q$. The actual number of individuals with character s_2 in the family should be $np = y'_{s_2}$, while the frequency for the $m - 1$ remaining classes within the family will be given by $y'_{s_1}, y'_{s_2}, y'_{s_3} \cdots y'_{s_m}$ providing (a) that the family is not differentiated from the population, *e. g.*, that there is no individual prepotency in the sense that we have used the term, and (b) that n is so large that the probable errors of random sampling are negligible. In actual work (b) can never, or almost never, be realized. Our problem is to determine whether differences between the theoretical class frequencies, y'_s , and the actually observed class frequencies, y''_s , in the family are to be regarded as due to chance merely or whether they are so large that they can reasonably be considered as indicating a differentiation of the family from the population to which it belongs. In short, our problem is to test $(y''_{s_2} - y'_{s_2})$ against its probable error.

Pearson has shown that the standard deviation of $(y''_s - y'_s)$ for any grade is

$$\Sigma(y''_s - y'_s) = \sqrt{npq \left(1 - \frac{n-1}{N-1}\right)}$$

and Tocher has pointed out that as a test for significance of divergence we may use either of the three ratios

$$(a) \quad (y_s'' - y_s') / \sqrt{npq(N - n) / (N - 1)}.$$

$$(b) \quad 100\{(y_s''/n) - p\} / \sqrt{100^2 p\bar{q}(N - n) / n(N - 1)}.$$

$$(c) \quad 100\{(y_s''/y_s') - 1\} / 100\sqrt{q(N - n) / np(N - 1)}.$$

The significance of these ratios can be judged from the tables of the probability integral.⁷

III. RECAPITULATION

Individual prepotency is here used to designate the superior capacity of certain parents for producing offspring of any desired character. The conception is most general, and does not imply a similarity in soma between parent and offspring, but the prepotency of the parent is judged entirely by the offspring it produces. The term is used merely to describe a long-known phenomenon, and no theoretical explanation is suggested.

Various breeders have tried to obtain a measure of individual prepotency in its present significance. The purpose of the present note is to point out certain biometric formulæ, in use for other purposes for several years, which seem well adapted for this purpose. They at least obviate several of the objectional features of some of the methods which have been employed. Their applicability in practical work will probably be limited by the arithmetical routine, but in experimental studies their importance may be very considerable. Illustration of their application will be published soon.

COLD SPRING HARBOR, N. Y.,

May 19, 1911.

⁷ Of course a statistical formula is not applicable to cases not covered by the assumptions on which it was developed. It seems unnecessary to discuss these here. Those using the formulæ should familiarize themselves with the limitations laid down by Pearson and Tocher in proposing the formulæ.